

Trabajo Fin de Grado

Salud auto percibida en España y sus
determinantes

Self-perceived health in Spain and it's
determinants

Autor

Carla Chavarrías Hernández

Director

María A. González Álvarez

Monia Ben-Kaabia Ben-Kaabia

Facultad de Economía y Empresa
Junio 2020

INDICE DE CONTENIDO

1.	RESUMEN	1
1.1.	PALABRAS CLAVE	1
2.	ABSTRACT	¡ERROR! MARCADOR NO DEFINIDO.
3.	INTRODUCCIÓN	2
4.	OBJETIVOS	3
5.	MARCO TEÓRICO	3
6.	METODOLOGÍA	7
6.1.	DESCRIPCIÓN DE LA BASE DE DATOS	7
6.2.	DESCRIPCIÓN DE LAS VARIABLES	8
6.3.	ANÁLISIS DE LAS VARIABLES	9
7.	MODELO DE ANÁLISIS	17
8.	APLICACIÓN ECONOMETRICA	23
8.1.	ESPECIFICACIÓN DEL MODELO	23
8.2.	ESTIMACIÓN DEL MODELO	24
8.3.	VALIDACIÓN DE LOS MODELOS ESTIMADOS	24
8.4.	CONTRASTES DE HIPÓTESIS	29
8.5.	BONDAD DE AJUSTE	30
8.6.	INTERPRETACIÓN DE LOS RESULTADOS	32
9.	CONCLUSIONES	38
10.	BIBLIOGRAFÍA	40

INDICE DE ILUSTRACIONES

Ilustración 1: Modelo de los determinantes de la salud de Dahlgren y Whitehead, 1991	4
Ilustración 2: Mapa de Autonomías según la significatividad de su relación con la buena salud auto percibida	14
Ilustración 3: Distribución logista	20
Ilustración 4: Función BMST de Artur Tarasso	25
Ilustración 5: Descripción de Gretl de la función BMST de Artur Tarasso	25
Ilustración 6: Detalles de la función BMST	26

INDICE DE TABLAS

Tabla 1: Porcentaje de salud buena y mala por edad	11
Tabla 2: Porcentajes de salud buena y mala según nivel de educación	11
Tabla 3: Porcentajes de salud buena y mala según el estado anímico	13
Tabla 4: Porcentaje de salud buena y mala por CCAA 1	15
Tabla 5: Porcentaje de salud buena y mala por CCAA 2	15
Tabla 6: Porcentaje de buena o mala salud según nacionalidad	16
Tabla 7: Porcentaje de buena o mala salud según estado civil	17
Tabla 8: Clasificación de los modelos de elección discreta	18
Tabla 9: Resumen de las variables utilizadas en el modelo	23
Tabla 10: Resultado de los contrastes para el modelo general	26
Tabla 11: Estimación del modelo Logit para toda la muestra (T=27714)	27
Tabla 12: Estimación del modelo Logit para los hombres (T=13218)	28
Tabla 13: Estimación del modelo Logit para las mujeres (T=14496)	29
Tabla 14: Comparación de los resultados entre hombres y mujeres	35
Tabla 15: Comparación de los ODD-Ratio y sus IC entre el modelo de hombres y el de mujeres	36

1. RESUMEN

El objetivo principal de este trabajo es realizar una investigación sobre los determinantes de la salud, entendiendo esta como salud auto percibida, teniendo en cuenta los factores: edad, sexo, estructura del hogar, nivel educativo, situación respecto a la actividad productiva, grado de urbanización de la zona de residencia, si se padecen o no enfermedades graves o crónicas, el grado de satisfacción con su vida y la frecuencia con la que se tienen tanto pensamientos felices como tristes o depresivos. Para realizar la investigación, se utilizarán los datos de la encuesta de condiciones de vida del año 2018 (instituto nacional de estadística).

En el marco del tema sobre el que se realiza este Trabajo de Fin de Grado (Herramientas Cuantitativas para el Análisis de Datos) se utiliza el software estadístico GRET, que permitirá desarrollar un modelo estadístico para determinar en qué medida cada uno de los factores considerados influye en el estado de salud subjetiva de los individuos.

1.1. PALABRAS CLAVE

Salud auto percibida, Encuesta de Condiciones de Vida INE 2018. Determinantes sociales de la salud, hombres, mujeres.

2. ABSTRACT

The main aim of this work is to carry out a research about the determinants of health, understanding this as self-perceived health, taking into account the factors: age, sex, household structure, educational level, situation regarding productive activity, degree of urbanization of the area in which they live, whether or not they suffer from serious or chronic diseases, degree of satisfaction with their life and the frequency with which they have both happy and sad or depressing thoughts.. This research is carried out using the results from the Survey on Living Conditions of the year 2018 (National Institute of Statistics).

In the framework of the subject in which this dissertation is conducted (Quantitative Tools for Data Analytics) we will use Gretl, as a statistics software, which will allow us to develop a statistical model that approximates the results we obtain to those of the article we use as a base.

3. INTRODUCCIÓN

La OMS (1948) definió la salud como “un estado de completo bienestar físico, mental y social, y no solamente la ausencia de afecciones o enfermedades”. Sin embargo, en la historia más reciente, el concepto que se tiene de salud es diferente, ya que no sólo engloba esa “ausencia de enfermedad”, sino que la entiende como un conjunto de factores personales, sociales, políticos, que determinan el nivel de salud tanto individual como poblacional. Analizar estos factores, puede ayudar a determinar diferencias entre grupos de población o incluso puede contribuir a crear perfiles que recojan las diferentes realidades de las personas, para que las entidades públicas puedan llevar a cabo acciones o políticas que reduzcan las desigualdades y mejoren la calidad de vida de la población.

En este trabajo, se pretende relacionar el estado de salud (gracias al concepto de salud auto percibida) de la población adulta en España, con algunos factores que la bibliografía en general relaciona con la salud. El término “salud auto percibida” es un valioso indicador para conocer la salud de la población, ya que recoge multitud de factores que el individuo, de manera consciente o inconsciente, relaciona con su salud a todos los niveles. Puede ser una herramienta muy útil para entender cómo afectan los Determinantes Sociales de la Salud (DSS) no sólo a la salud de las personas, sino también a su percepción de esta.

Se parte de las hipótesis de que los DSS explican el nivel de salud de los individuos, así como de la capacidad de la salud auto percibida para ilustrar la realidad en lo que a salud se refiere de los individuos. Los factores que analizaremos son la edad, el sexo, el nivel educativo y la estructura del hogar, que son indicadores que se pueden asociar directamente a los DSS. Para conseguir los objetivos formulados, se utilizará

una metodología cuantitativa con la que se analizarán los resultados de la Encuesta sobre Condiciones de Vida del INE en el año 2018.

El presente trabajo se expone un marco teórico, unos objetivos y una metodología que ayudarán a contextualizar y comprender los resultados obtenidos. En primer lugar, se llevará a cabo un modelo general, que englobe todos los segmentos de la muestra para desglosarlo en los siguientes modelos según sexo.

4. OBJETIVOS

El objetivo general de la investigación es analizar el nivel de salud auto percibida de la población adulta en España, así como sus principales determinantes.

Para un estudio en profundidad, se establecen también los siguientes objetivos específicos:

- A. Conocer el nivel de salud auto percibida de la población española en general y atendiendo a las diferencias por sexo.
- B. Identificar los principales determinantes que influyen en la salud auto percibida de la población española.
- C. Analizar la relación entre la salud auto percibida y los factores socioculturales a través del nivel educativo.
- D. Relacionar la estructura del hogar con la valoración de la salud auto percibida.
- E. Identificar si existen diferencias significativas en la salud auto percibida según criterios de género y edad.

5. MARCO TEÓRICO

El concepto de salud ha evolucionado a lo largo de los años. En la constitución de la OMS en 1948, se relacionaba la salud con la ausencia de enfermedades. Sin embargo, hoy en día, pasamos a entenderla como un conjunto de factores que vienen determinados por la biología, los estilos de vida, las relaciones del individuo y las condiciones socioeconómicas, culturales y ambientales a las que se ve sometido.

Según la OMS, “el goce del grado máximo de salud que se pueda lograr es uno de los derechos fundamentales de todo ser humano.” Es por ello, que los gobiernos y entidades públicas deben afanarse por reducir las desigualdades en lo referente a la salud, porque hoy en día un buen estado de salud sigue estando ligado a la posición socioeconómica de los ciudadanos (Dahlgren y Whitehead 2006).

A la hora de planificar políticas sanitarias, es imprescindible conocer las necesidades de la población a este respecto ya que no todas las personas tienen las mismas necesidades. En 1998 la OMS definió los *determinantes de la salud* como “un conjunto de factores personales, sociales, políticos y ambientales que determinan el nivel o estado de salud de los individuos y las poblaciones”. En esta definición, la OMS quiso englobar el bienestar biológico, el psicológico y el social.

En 1991, Dahlgren y Whitehead desarrollaron un modelo “Rainbow Model” que explicaba los determinantes de la salud como “capas de influencia” (Frenz, 2005) que van interactuando con el individuo. Esta explicación se entiende fácilmente si observamos la Figura1.

Ilustración 1: Modelo de los determinantes de la salud de Dahlgren y Whitehead, 1991



En este modelo, podemos ver, por un lado, los factores como el sexo, la edad, o las características constitucionales, ninguno de ellos se considera modificable en este estudio. Además, resalta especialmente lo que denominamos “Determinantes Sociales de la Salud”, que teóricamente, son susceptibles de ser modificados. Estos últimos, se dividen en tres niveles: el primero hace referencia al estilo de vida del individuo y las conductas relacionadas con la salud que desarrolla, el segundo hace referencia a las relaciones sociales y las interacciones con diversas personas, el tercer y último nivel,

tiene que ver con las condiciones socioeconómicas, culturales y ambientales en las que se ve envuelto el individuo.

Tanto en nuestro estudio, como en el que se usa como referencia, se tendrán en cuenta variables que hacen referencia a las características inherentes a cada persona (edad y sexo), así como a factores sociales (tipo de hogar) y a factores socioculturales (nivel educativo).

Es de vital importancia el concepto de “*salud auto percibida*”, que se podría definir como “percepción del individuo sobre su estado de salud general”. Esta puede aproximar bien el nivel de salud general de los individuos, ya que, a pesar de ser subjetivo, su simplicidad y perfil multidimensional, han hecho que la OMS lo recomiende como estándar de las encuestas de salud (Sánchez, 2017). La medición de este parámetro suele darse en encuestas por medio de la pregunta “¿Cómo valora Ud. el estado de su salud en general?”. Las posibles respuestas a la misma tienden a darse mediante una escala de Likert de 5 puntos, en el que 1 hace referencia a “Muy bueno”, 2 a “bueno”, 3 “regular”, 4 “malo” y 5 “muy malo”. En muchos estudios se utiliza este valor como si se tratase de una variable objetiva “estado de salud general”, ya que ésta última sería muy compleja de medir, hay múltiples factores que intervienen en ella y sería muy complicado estimar un rango de valores objetivos. Además, la salud auto percibida, aporta también factores subjetivos del individuo, lo que puede llegar a enriquecer los estudios. Es por ello que, para valorar la salud de la población objeto de estudio nos valdremos de este concepto que, además, es el que utiliza el INE en su Encuesta de Condiciones de Vida.

Encontramos diferentes estudios que relacionan la salud auto percibida con distintos estratos de la población según edad, diferenciando principalmente tres grupos: “niños y adolescentes”, “adultos” y “mayores de 60 o 65 años”. Partimos de la base de que la encuesta sobre la que trabajamos, sólo se realiza a personas mayores de edad, por lo que el primer grupo quedaría automáticamente fuera del estudio. En lo que hemos llamado “adultos”, encontramos diferente bibliografía que considera dentro del grupo a mayores de 25 o 30 años, por diferentes motivos. Al existir tantas discrepancias en la concepción de grupos de edad en estudios sobre este tema, nosotros tomaremos la

distribución de la muestra como punto de partida para determinar los grupos de edad con los que trabajaremos.

Existen trabajos (Aguilar-Palacio et al., 2014) que corroboraron la relación entre el nivel educativo y la salud percibida, hasta el punto de considerar al primero como un buen estimador de la salud subjetiva. Las personas con niveles educativos más bajos muestran en general peores niveles de percepción de su salud. También es este grupo de población quien presenta peores indicadores de estilo de vida (mayor consumo de tabaco y alcohol, niveles mayores de sedentarismo, o problemas para llevar a cabo las actividades diarias). Esta tendencia, se ve de manera más dramática en el caso de las mujeres, de manera que podemos destacar la importancia que tiene la educación y, por tanto, de los trabajos que desempeñan, en la salud de éstas (Rohlf et al., 2000).

Entendemos el hogar como el conjunto de las personas con las que se convive, estableciendo intercambios continuos de diferentes recursos (Gomà et al., 2017). Las relaciones que se dan en el seno del hogar y la estructura de este resultan de gran importancia a la hora de valorar la calidad de vida relacionada con la salud. (Gomà et al., 2015). Las variables más importantes dentro del hogar son: la convivencia o no en pareja (que en muchos casos simplificamos identificando como matrimonio) la tenencia o no de hijos, si se trata o no de un hogar monoparental y, por supuesto, la importancia de los roles de género en esta estructura de convivencia.

En lo referente al sexo, Morcillo (2017) recoge una amplia bibliografía que relaciona el sexo femenino con una peor salud auto percibida. Sin embargo, encontramos dos estudios que no defienden esta relación, ya que, o bien atribuyen esta relación al entorno cultural de la investigación (Artacoz, 2004), o bien encuentran esta relación, pero no la consideran causal, sino que piensan que se da por otros factores estadísticos (Girón, 2010). Si bien es cierto que ambos estudios se reducen a un ámbito territorial muy delimitado, pudiendo ser esta una de las causas de sus conclusiones. Además, se dan otras dos condiciones relacionadas con la salud auto percibida en mujeres: en primer lugar, la *Paradoja de la mortalidad* que recoge el contraste que supone la mayor esperanza de vida de las mujeres, en contraposición su percepción de la salud propia. Según Morcillo, esto podría deberse a que las mujeres sufren más afectaciones crónicas, que empeoran su calidad de vida, pero que están menos

relacionadas con la mortalidad de lo que están las que padecen los hombres. Esta no es la única explicación que encontramos a este fenómeno, desde Verbrugge (1989), son numerosos los estudios que relacionan estos niveles peores de salud auto percibida con factores sociales, en especial con aquellos relacionados con su situación laboral y con su dedicación al trabajo doméstico (aparecen en Morcillo, 2017).

6. METODOLOGÍA

Este estudio se estiman modelos que ilustran la relación existente entre la salud auto percibida y diferentes DSS partiendo de los datos proporcionados por la Encuesta de Condiciones de Vida del Instituto Nacional de Estadística de 2018. En él se aplica metodología cuantitativa, realizando análisis estadísticos mediante la utilización del software Gretl. La recogida de datos de esta encuesta se llevó a cabo durante el tercer cuatrimestre de 2018 mediante una entrevista personal que realizada a alrededor de 13.000 hogares y un total de 53.000 personas.

La Encuesta de Condiciones de Vida se ha utilizado en diferentes estudios científicos como instrumento estadístico para analizar la pobreza, desigualdad, cohesión social, necesidades de la población, impacto de las políticas sociales y económicas y diseño de nuevas políticas tanto en el ámbito nacional como en el europeo, ya que utiliza criterios armonizados en todos los países miembros.

6.1. DESCRIPCIÓN DE LA BASE DE DATOS

Los datos recogidos se organizan en varios cuestionarios separados: ficha de seguimiento, ficha de hogar, cuestionario de hogar y cuestionario individual. Dentro del cuestionario sobre el hogar, existen módulos temáticos sobre aspectos como la renta, la exclusión social, o la vivienda. Y en el cuestionario individual encontramos módulos específicos sobre educación, información laboral, salud, etc.

Cabe destacar que esta encuesta es una encuesta panel, en la que las personas entrevistadas participan durante cuatro años seguidos, así como que los datos sobre la renta hacen referencia al año anterior al de la recogida de datos, en este caso, 2017.

La muestra efectiva para el estudio está formada por unas 34.000 personas que tras un primer filtrado de la base de datos se ha reducido a 27.714 casos válidos para este análisis.

6.2. DESCRIPCIÓN DE LAS VARIABLES

Para estudiar las variables que condicionan la salud auto percibida de los españoles en 2018, se decide escoger aquellas variables que condicionan de manera fundamental el estilo de vida y el entorno de las personas y que, por lo tanto, se pueden considerar DSS. Es decir: el sexo, la edad, el nivel de estudios y el tipo de hogar en el que se viva. Asimismo, también se ha decidido incluir en el estudio otros factores sociales, socioeconómicos y socioculturales que podrían influir tanto en la salud como en las percepciones de los individuos: el grado de urbanización del lugar donde viven, si padecen o no enfermedades crónicas, el grado de satisfacción con su vida, la frecuencia con la que tienen sentimientos negativos o depresivos y también felices, su situación respecto a la actividad productiva y la región en la que viven.

De este modo, las variables independientes que se utilizan en el estudio son las siguientes:

- Sexo: variable dicotómica que distingue entre hombres y mujeres.
- Edad: es una variable continua que toma valores entre 17 y 86 años. Para algunos análisis se ha recodificado en tres categorías: jóvenes de 17 a 30, edad media de 31 a 65 y mayores de más de 65 años.
- Nivel de estudios: Esta variable presentaba seis categorías más cuatro subcategorías, por lo que ha sido reclasificada en tres categorías: estudios primarios, estudios secundarios y estudios superiores (cada categoría comprende los estudios cursados y los finalizados de mayor nivel).
- Tipo de hogar: recodificada en hogares monoparentales, hogares sin hijos dependientes y hogares con hijos dependientes (se consideran hijos dependientes los menores de edad y los que tienen 18 años o más, pero menos de 25 y son económicamente inactivos).

- Grado de urbanización: distingue dos categorías: zona urbana, que engloba las muy pobladas y las zonas de densidad de población media y zona rural, que incluye a las poco pobladas.
- Enfermedad: Hace referencia a si el encuestado padece o no enfermedades graves o crónicas. Se distingue entre quienes sí las sufren y quiénes no.
- Grado de satisfacción con su vida: Está reflejada en una escala de Likert de 10 puntos de menor a mayor grado de satisfacción con la vida del encuestado. Esta escala será considerada como una variable continua.
- Frecuencia de sentimientos o pensamientos tristes o depresivos: Ofrece cinco respuestas posibles: siempre, casi siempre, algunas veces, sólo alguna vez, y nunca.
- Frecuencia de pensamientos o sentimientos felices: distingue las mismas cinco categorías que la variable anterior: siempre, casi siempre, algunas veces, sólo alguna vez, y nunca.
- Situación respecto a la actividad productiva: Sólo se distinguen dos categorías, parado o activo.
- Región: Se mantienen las diecinueve categorías de la variable original, que incluyen las diecisiete comunidades autónomas y las dos ciudades autónomas del territorio español.
- Nacionalidad: Se distingue entre extranjeros y de nacionalidad europea.
- Estado civil: Solamente se distingue entre solteros y casados, el caso de los divorciados y viudos se contabilizará como soltero y el de los separados como casados.

Como variable dependiente se ha utilizado la salud auto percibida de los encuestados que en los datos originales aparecía reflejada en cinco categorías (muy buena, buena, regular, mala y muy mala), pero que se ha recodificado dando únicamente dos valores: buena (que toma valor uno para aquellos individuos cono buena o muy buena salud auto percibida) y mala (que toma valor cero para el resto de las categorías).

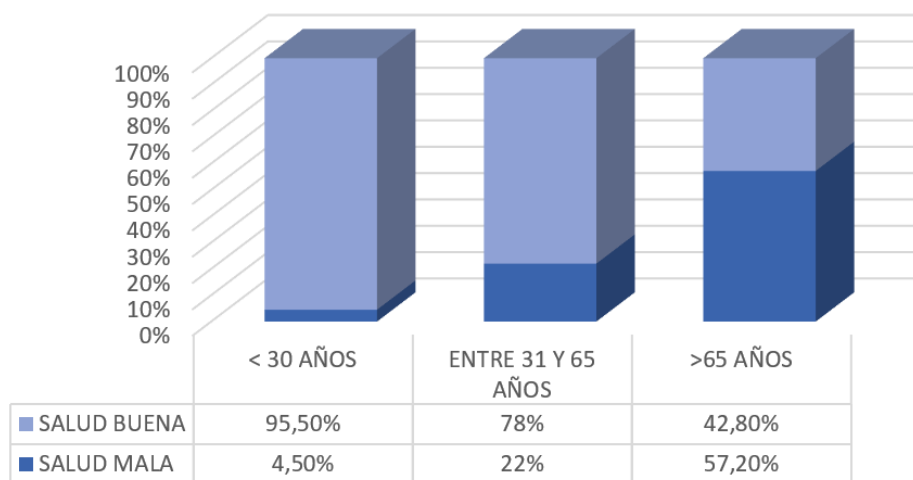
6.3. ANÁLISIS DE LAS VARIABLES

Antes de modelizar, se analiza cada variable de manera descriptiva, binomial y, en algunos casos, se recodifica la misma de acuerdo con los resultados de dicho análisis. Con ello se persigue conocer de una manera más profunda la estructura de la muestra y de las variables, así como describir posibles relaciones entre éstas.

Los datos indican que en 2018 el 71,86% de la población consideraba tener buena salud, frente al 28,14% que consideraba su salud mala. A continuación, se describen las principales conclusiones obtenidas de este análisis descriptivo:

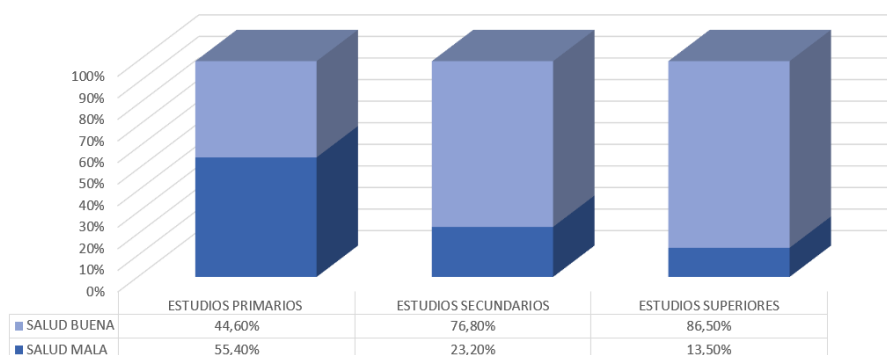
- Sexo: teniendo en cuenta toda la población, los porcentajes de respuesta de hombres y mujeres son similares (47,7% y 52,3% respectivamente), además, el contraste chi-cuadrado de Pearson confirma que existe asociación entre esta variable y la salud auto percibida. En el análisis bivariante se observa cómo es mayor el porcentaje de hombres que refieren buena salud (74,7%) que el de mujeres (69,3%). (Véase *¡Error! No se encuentra el origen de la referencia.*).
- Edad: Como se puede observar en las *¡Error! No se encuentra el origen de la referencia.* y *¡Error! No se encuentra el origen de la referencia.* del Anexo, la media de edad poblacional (aproximando, 52 años) es mayor que la media de edad de las personas que dicen tener buena salud (aproximando 47 años). Además, la hipótesis de que la edad es un factor que influye en la salud auto percibida se confirma mediante los resultados que arroja el test de contraste de muestras (Véanse *¡Error! No se encuentra el origen de la referencia.* y *REF_Ref42884483 \h * MERGEFORMAT ¡Error! No se encuentra el origen de la referencia.* del Anexo). Como se puede ver en el siguiente gráfico, las respuestas que hacen referencia a buena salud representan un porcentaje más alto cuanto más baja es la edad.

Tabla 1: Porcentaje de salud buena y mala por edad



- Nivel de estudios: Casi la mitad de los encuestados tienen estudios secundarios, por lo que será la categoría con mayor representación en el estudio (Véase **¡Error! No se encuentra el origen de la referencia.** del Anexo). La relación entre esta variable y la variable *buena salud* es significativa, tal y como refleja el contraste chi-cuadrado ($p=0$). La relación entre la buena salud auto percibida y las diferentes categorías de nivel de estudios se muestra en el Gráfico 2 a continuación:

Tabla 2: Porcentajes de salud buena y mala según nivel de educación



Se observa que: a mayor nivel de estudios, mejor nivel de salud reportan los encuestados, es decir, el coeficiente de correlación entre estas dos variables es positivo.

Tipo de hogar: Esta variable, que originalmente tenía catorce categorías, se han recodificado para dar lugar únicamente a tres (hogares unipersonales,

hogares sin niños dependientes y hogares con niños dependientes). Este cambio hace que la distribución de frecuencias de las categorías con las que se trabajará sea la que se observa en el Anexo, en la Tabla 9. El p-valor del contraste chi-cuadrado confirma lo que podemos intuir en la siguiente tabla: que la relación entre las variables es significativa, siendo los hogares constituidos por una única persona los que reportan peor salud en mayor medida.

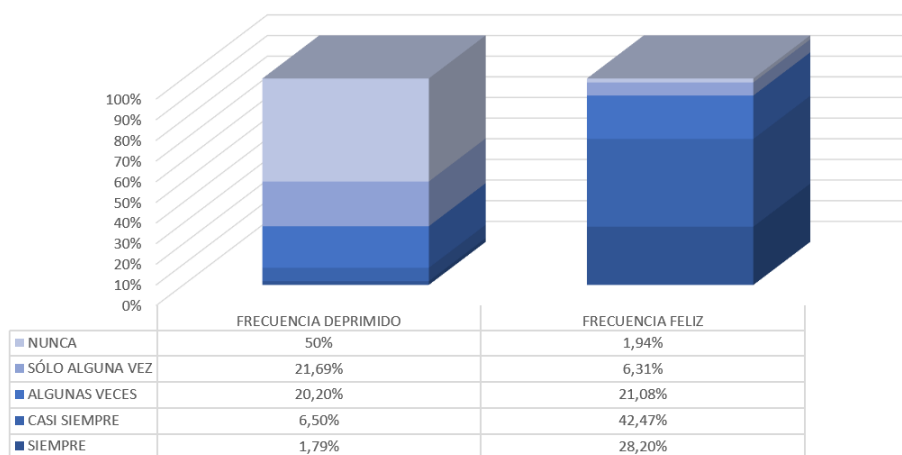
- Grado de urbanización: La conclusión a la que se llega después de analizar las interacciones entre esta variable y la endógena en cada una de las tres categorías con las que se trabaja (zona muy poblada, media y poco poblada) es que las zonas con una densidad de población media refieren un porcentaje más alto de casos de buena salud y que las zonas menos pobladas tienen un mayor porcentaje de respuestas que indican mala salud. Esta última relación podría deberse a que estas zonas menos pobladas se refieren principalmente a entornos rurales donde la edad media es muy superior a la de los entornos urbanos. Pero estas posibles relaciones serán captadas por el análisis multivariante que se expone en el siguiente apartado (Véase Tabla 22 del Anexo). Aunque estas diferencias nos son importantes, debido al valor que toma el p-valor, podemos afirmar que la variable *grado de urbanización* puede ser relevante a la hora de determinar la salud auto percibida.
- Enfermedad: Un 35,11% de la muestra dice sufrir algún tipo de enfermedad crónica o grave, lo que sin duda influye en el nivel general de salud que percibe tener de una manera significativa ($p=0$) (Véase Tabla 23 del Anexo).
- Grado de satisfacción con su vida: Como podemos ver en la Tabla 24 y la Tabla 25 del Anexo, la media prácticamente no sufre variaciones si tenemos en cuenta toda la muestra o si sólo tenemos en cuenta a quienes refieren buena salud, sin embargo, según el estadístico de contraste, esta variable sí es significativa respecto del nivel de salud. Esta relación se puede achacar a que el grado de satisfacción con la vida puede significar un mayor optimismo y que esta manera de ver la vida influye directamente también en la percepción sobre la salud.
- Frecuencia de sentimientos o pensamientos tristes o depresivos: Esta variable y la percepción de tener buena salud están relacionadas de manera inversa, cuanto más comunes son estos sentimientos de desánimo, más comunes son las

respuestas que indican mala salud. El coeficiente de correlación de estas variables es positivo (debido al orden de las categorías de la variable que van de más a menos frecuente). En las 27 y 28, podemos ver los datos relativos al contraste de medias de esta variable que se ha calculado respecto de la media poblacional y la media de la población que refiere buena salud, el test permite rechazar la hipótesis nula de igualdad de medias. En Tabla 26, podemos observar cómo es mucho más frecuente no tener pensamientos depresivos que tenerlos con asiduidad.

- Frecuencia de pensamientos o sentimientos felices: Existe una notable diferencia entre la media poblacional (4,1161) y la media de quienes presentan buena salud (4,3479). Esta diferencia nos indica que la variable que indica la frecuencia con la que se tienen pensamientos felices influye notablemente sobre el grado de salud auto percibida. Una posible explicación podría girar en torno a la hipótesis de que un mayor optimismo implica también una visión más optimista de la salud, o que la ausencia de grandes preocupaciones repetidas en el tiempo (o la alta frecuencia de exposición a sentimientos positivos) llega a mejorar la percepción de la salud o bien la salud en sí misma.

En la siguiente gráfica podemos observar de una manera más clara la influencia del estado anímico en la salud de los encuestados.

Tabla 3: Porcentajes de salud buena y mala según el estado anímico

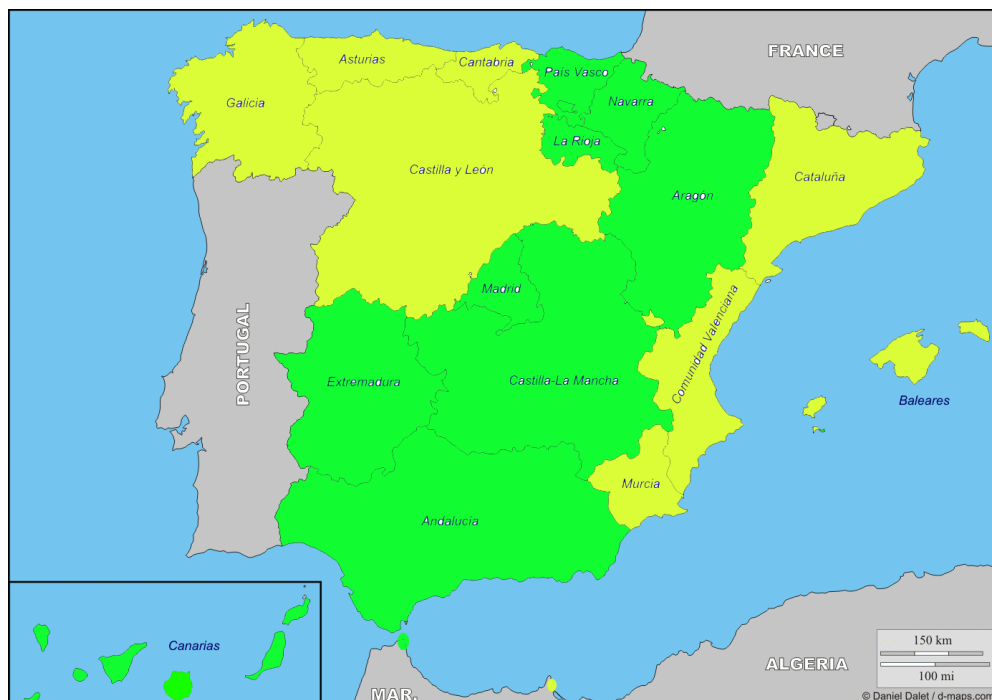


- Situación respecto a la actividad productiva: Según el contraste chi-cuadrado ($p=0$), las variables actividad y buena salud están asociadas. Quienes gozan de mejor salud son las personas que están trabajando, seguidos por los parados y

quienes peor salud tienen son los jubilados, esto puede deberse al factor edad, ya que los jubilados suelen ser más mayores que las personas que están trabajando y, como hemos visto anteriormente, estas personas refieren peor salud. (Véase Tabla 29 del Anexo).

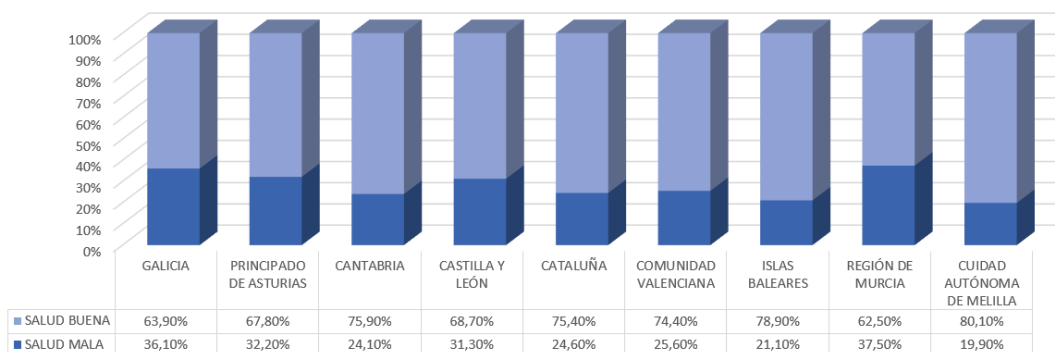
- **Región:** Para analizar esta variable, se ha convertido a todas las categorías en variables dependientes dicotómicas, viendo también la interacción que encontrábamos entre éstas y la variable endógena. Tras hacerlo, se observa cómo algunas regiones presentan una relación significativa con la buena salud auto percibida, pero otras no. En la siguiente imagen se pueden distinguir en color amarillo las autonomías cuya relación con la variable dependiente es significativa y en verde, aquellas cuya relación no es significativa.

Ilustración 2: Mapa de Autonomías según la significatividad de su relación con la buena salud auto percibida



El siguiente gráfico representa el porcentaje de salud buena y mala que refleja cada una de las comunidades que sí está asociada con una buena salud auto percibida según el contraste chi-cuadrado.

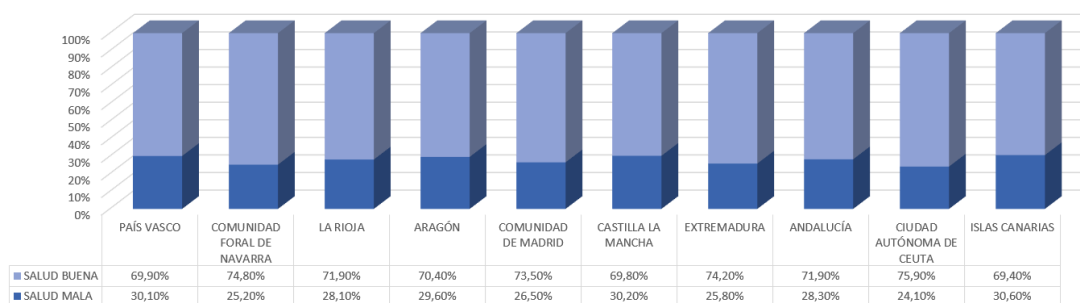
Tabla 4: Porcentaje de salud buena y mala por CCAA 1



Como podemos ver, hay dos valores que destacan por encima de los demás: Galicia y Región de Murcia, que presentan valores más altos de mala salud que el resto de las comunidades. En el sentido contrario, destacan las Islas Baleares y la Ciudad Autónoma de Melilla, que reflejan mayores porcentajes de buena salud que el resto. Aunque también es verdad, que estas dos últimas son de las comunidades que menos respuestas ofrecen. Esto podría influir en el resultado que se ha obtenido.

El siguiente gráfico representa el porcentaje de salud buena y mala que refleja cada una de las comunidades que no está asociada con una buena salud auto percibida según el contraste chi-cuadrado.

Tabla 5: Porcentaje de salud buena y mala por CCAA 2



Los valores en todas estas comunidades son muy similares, situándose casi todos ellos alrededor del 70-75% de población con buena salud. Sin embargo, hay un dato que es ligeramente más bajo que el resto. Es el caso de la Ciudad Autónoma de Ceuta, que refiere mejores valores porcentuales de buena salud, sin embargo, en esta ciudad se han realizado un porcentaje muy bajo del total de las encuestas realizadas y esta respuesta

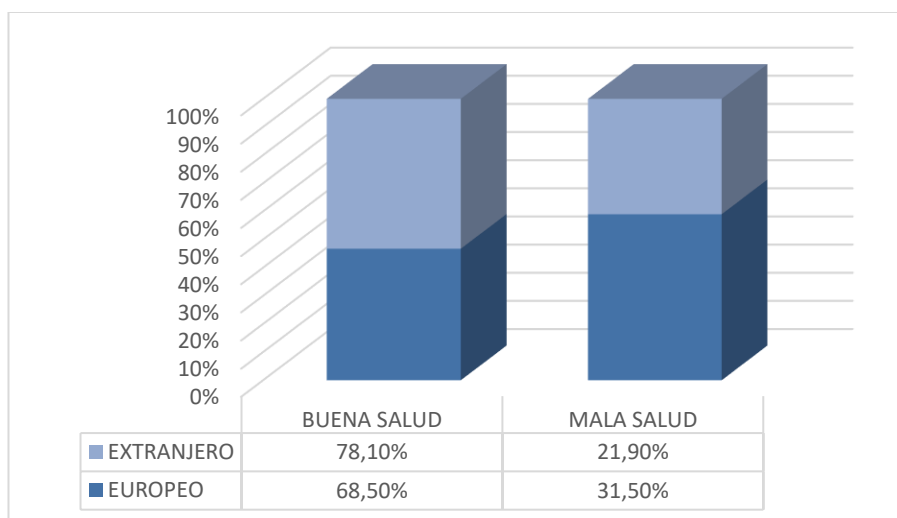
tan baja puede alienar la representatividad de los datos frente a los ofrecidos por otras regiones.

Debido a que en algunas de las comunidades, la diferencia de medias es significativa y tienen una proporción de personas que perciben su salud como buena por encima de la media (como Cantabria y Baleares), otras con una mayor proporción de salud auto percibida mala (como Murcia, Galicia, Asturias o Castilla y León) y, por último, hay una serie de comunidades autónomas que se sitúan en torno a la media poblacional y para las que el contraste de medias no es significativo, no se incluirá esta variable en el modelo.

- Nacionalidad: Esta variable, originalmente tenía tres categorías: nacionalidad española, nacionalidad europea diferente a la española y extranjero de terceros países, sin embargo, se ha modificado para que solamente tome dos categorías: europeo (entendiendo a las personas de nacionalidad española dentro de este factor) y extranjero, que tiene en cuenta sólo a personas con nacionalidad de países que no formen parte de la UE.

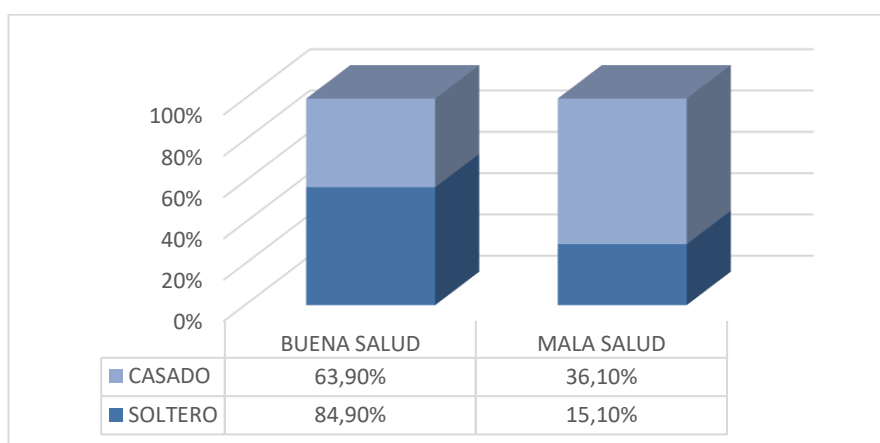
Si se tiene en cuenta la buena o mala salud, podemos ver en el siguiente gráfico que el porcentaje de personas que refieren mala salud es mayor en extranjeros que en europeos.

Tabla 6: Porcentaje de buena o mala salud según nacionalidad



- Estado civil: Originalmente, esta variable tenía cinco categorías: Casado, soltero, divorciado, separado y viudo. Se ha decidido simplificar la variable atendiendo únicamente al criterio de si son solteros o casados, incluyendo en el primer grupo a viudos y divorciados y en el segundo a separados (ya que en su estado civil “oficial” sigue constando como casado).

Tabla 7: Porcentaje de buena o mala salud según estado civil



Como se puede ver en la gráfica anterior, el porcentaje de personas que refieren mala salud duplica en el caso de los casados al valor que toma en los solteros, por lo que, *a priori*, se ve una relación clara que relaciona una peor salud con los casados.

7. MODELO DE ANÁLISIS

El objetivo del análisis llevado a cabo en el presente estudio es conocer la influencia de las variables tanto cuantitativas como cualitativas sobre la probabilidad de que un individuo tenga una buena salud auto percibida. Para la realización del análisis se hará uso de un modelo de elección discreta a través del uso de variables ficticias, formadas por un número finito de alternativas.

La diferencia de los modelos de elección discreta frente a la econometría tradicional radica en que los primeros permiten la modelización de variables cualitativas, a través del uso de técnicas propias de las variables discretas. Se dice que una variable es discreta cuando está formada por un número finito de alternativas que miden cualidades. Esta característica exige la codificación como paso previo a la modelización, proceso por el cual las alternativas de las variables se transforman en

códigos o valores cuánticos, susceptibles de ser modelizados utilizando técnicas econométricas.

La modelización de este tipo de variables se conoce genéricamente con el nombre de modelos de elección discreta, dentro de la cual existe una amplia tipología de modelos. En concreto, según el número de alternativas incluidas en la variable endógena, se distinguen los modelos de respuesta dicotómica frente a los denominados modelos de respuesta o elección múltiple. Según la función utilizada para la estimación de la probabilidad existe el modelo de probabilidad lineal truncado, el modelo Logit y el modelo Probit. Según que las alternativas de la variable endógena sean excluyentes o incorporen información ordinal se distingue entre los modelos con datos no ordenados y los modelos con datos ordenados. Dentro de los primeros, según que los regresores hagan referencia a aspectos específicos de la muestra o de las alternativas entre las que se ha de elegir, se distingue entre los modelos multinomiales y los condicionales.

Teniendo en cuenta todos los elementos que influyen en el proceso de especificación de los modelos de elección discreta, se puede establecer una clasificación general de los mismos, que queda recogida en la Tabla 8.

Tabla 8: Clasificación de los modelos de elección discreta

Nº de alternativas	Tipo de alternativas	Tipo de función	El regresor se refiere a:	
			Características	Atributos
			(de los individuos)	(de las alternativas)
Modelos de respuesta dicotómica (2 alternativas)	Complementarias	Lineal	Modelo de Probabilidad Lineal Truncado	
		Logística	Modelo Logit	
		Normal tipificada	Modelo Probit	
Modelos de respuesta múltiple (más de 2 alternativas)	No ordenadas		Logit Multinomial	Logit Condicional
		Logística	- Logit Anidado	- Logit Anidado
			- Logit Mixto	- Logit Mixto
		Normal tipificada	Probit Multinomial	Probit Condicional
			Probit	Probit

			Multivariante	Multivariante
	Ordenadas	Logística	Logit Ordenado	
		Normal tipificada	Probit Ordenado	

Dado que el uso de una función de distribución garantiza que el resultado de la estimación esté acotado entre 0 y 1, en principio las posibles alternativas son varias, siendo las más habituales la función de distribución logística, que ha dado lugar al modelo Logit, y la función de distribución de la normal tipificada, que ha dado lugar al modelo Probit. Tanto los modelos Logit como los Probit relacionan, por tanto, la variable endógena Y_i con las variables explicativas X_{ki} a través de una función de distribución.

En el caso del modelo Logit, la función utilizada es la logística, por lo que la especificación de este tipo de modelos queda como sigue.

$$\text{LOGIT}; Y_i = \frac{1}{1 + e^{-\alpha - \beta_k X_{ki}}} + \varepsilon_i = \frac{e^{\alpha + \beta_k X_{ki}}}{1 + e^{\alpha + \beta_k X_{ki}}} + \varepsilon_i \quad (1)$$

En el caso del modelo Probit la función de distribución utilizada es la de la normal tipificada, con lo que el modelo queda especificado a través de la siguiente expresión,

$$\text{PROBIT}; Y_i = \int_{-\infty}^{\alpha + \beta X_i} \frac{1}{(2\pi)^{1/2}} e^{-\frac{s^2}{2}} ds + \varepsilon_i \quad (2)$$

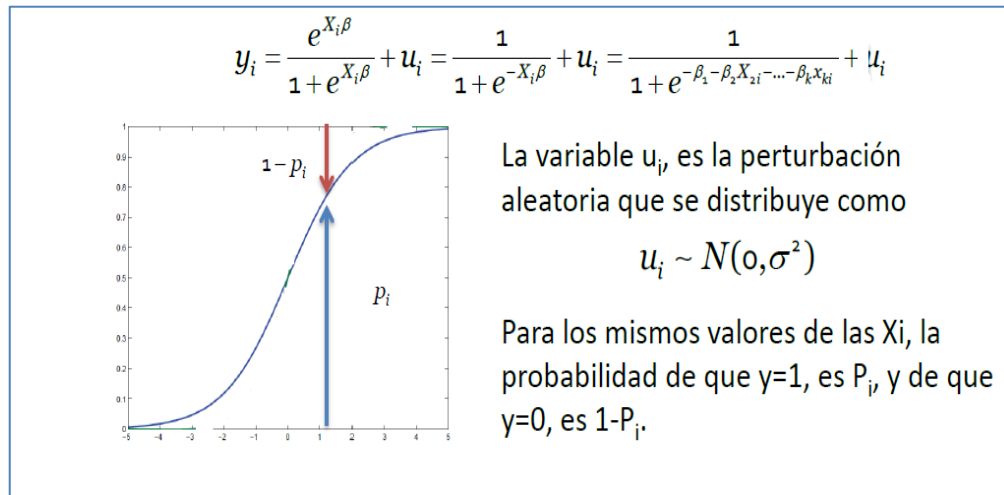
donde la variable s es una variable “muda” de integración con media cero y varianza 1.

Dada la similitud existente entre las curvas de la normal tipificada y de la logística, los resultados estimados por ambos modelos no difieren mucho entre sí, siendo las diferencias operativas, debidas a la complejidad que presenta el cálculo de la función de distribución normal frente a la logística, ya que la primera solo puede calcularse en forma de integral. La menor complejidad de manejo que caracteriza al modelo Logit es lo que ha potenciado su aplicación en la mayoría de los estudios empíricos.

Al igual que en el Modelo Lineal de Probabilidad, el Modelo Logit (1) se puede

interpretar en términos probabilísticos, es decir, sirve para medir la probabilidad de que ocurra el acontecimiento objeto de estudio ($Y_i=1$).

Ilustración 3: Distribución logista



Partiendo de la ecuación general del Modelo Logit (2.8) y definido P_i como la probabilidad del estado o la alternativa 1, y dados los valores de las variables explicativas x_2, x_3, \dots, x_k , las probabilidades de que la variable dependiente tome los valores 1 y 0 son:

$$P_i = \text{Pr ob}(Y_i = 1 | x_2, \dots, x_k) = E(Y_i | X = x_i) = \frac{e^{\beta_1 + \beta_2 x_{2i} + \dots + \beta_k x_{ki}}}{1 + e^{\beta_1 + \beta_2 x_{2i} + \dots + \beta_k x_{ki}}} \quad (3)$$

$$\begin{aligned} \text{Pr ob}(Y_i = 0 | x_2, \dots, x_k) &= 1 - P_i \\ &= 1 - \frac{e^{\beta_1 + \beta_2 x_{2i} + \dots + \beta_k x_{ki}}}{1 + e^{\beta_1 + \beta_2 x_{2i} + \dots + \beta_k x_{ki}}} \\ &= \frac{1}{1 + e^{\beta_1 + \beta_2 x_{2i} + \dots + \beta_k x_{ki}}} \end{aligned} \quad (4)$$

Dada una muestra poblacional y variable aleatoria caracterizada por unos parámetros, se consideran estimadores Máximo-Verosímiles de los parámetros de una población determinada, aquellos valores para los cuales la función de densidad conjunta (o función de verosimilitud) alcanza un máximo. El estimador máximo verosímil del modelo Logit es consistente, asintóticamente eficiente y asintóticamente normal.

En cuanto a la interpretación de los parámetros estimados en un modelo Logit, el signo de los mismos indica la dirección en que se mueve la probabilidad cuando aumenta la variable explicativa correspondiente, sin embargo, la cuantía del parámetro no coincide con la magnitud de la variación en la probabilidad (como si ocurría en el MLP). En el caso de los modelos Logit, al suponer una relación no lineal entre las variables explicativas y la probabilidad de ocurrencia del acontecimiento, cuando aumenta en una unidad la variable explicativa los incrementos en la probabilidad no son siempre iguales ya que dependen del nivel original de la misma.

Partiendo de la expresión del modelo LOGIT, el efecto marginal de una variable sobre la probabilidad de que ocurra el suceso (ser pobre) se obtiene calculando la derivada parcial con respecto a las explicativas:

$$\frac{\partial P(Y_i = 1)}{\partial X_{ji}} = \frac{e^{(\beta_1 + \beta_2 X_{2i} + \dots + \beta_k X_{ki})}}{(1 + e^{(\beta_1 + \beta_2 X_{2i} + \dots + \beta_k X_{ki})})^2} \times \beta_j \quad (5)$$

Por tanto, el efecto marginal en ambos modelos depende de los valores que toman las variables explicativas, lo que hace que el efecto marginal de una variable explicativa sobre la probabilidad no es constante. Pueden calcularse los efectos marginales para cada observación de la muestra. Para evitar esto, se suelen calcular los efectos marginales para los valores medios de las variables explicativas:

$$\frac{\partial P(Y_i = 1)}{\partial X_{ji}} = \frac{e^{(\beta_1 + \beta_2 \bar{X}_{2i} + \dots + \beta_k \bar{X}_{ki})}}{(1 + e^{(\beta_1 + \beta_2 \bar{X}_{2i} + \dots + \beta_k \bar{X}_{ki})})^2} \times \beta_j = P_{X=\bar{X}}(1 - P_{X=\bar{X}}) \times \beta_j \quad (6)$$

$P(1 - P) \times \beta_j \rightarrow$ "pendiente"

Cuando se estima un modelo Logit en Gretl, en la pantalla principal hay una columna con el valor de los coeficientes estimados, otra con el valor de sus desviaciones típicas estimadas, otra con el valor del t-ratio, y otra cuarta que lleva el nombre de "pendientes". Esas pendientes son precisamente el producto del parámetro estimado multiplicado por el valor de la función de densidad valorado en el valor promedio de las variables exógenas. Pretende recoger el incremento en la probabilidad de un individuo medio al incrementar en una unidad una determinada variable explicativa X_k .

Cuando una variable explicativa es binaria (FICTICIA 0 y 1) el efecto de que $x_k=0$ a $x_k=1$ se calcula como diferencia de probabilidades:

$$\hat{P}_{x_k=1} - \hat{P}_{x_k=0} \quad (7)$$

Una interpretación más sencilla de los parámetros estimados es la que se obtiene a través de la linealización del modelo. Para ello, partiendo de la expresión general del modelo LOGIT:

$$\frac{\hat{P}_i}{1 - \hat{P}_i} = e^{\hat{\beta}_1 + \hat{\beta}_2 X_{2i} + \dots + \hat{\beta}_k X_{ki}}$$
$$ODD = \frac{\hat{P}_i}{1 - \hat{P}_i} = e^{\hat{\beta}_1 + \hat{\beta}_2 X_{2i} + \dots + \hat{\beta}_k X_{ki}} \quad (8)$$

- Al cociente entre la probabilidad de que ocurra un hecho, o de que se elija la opción 1, frente a la probabilidad de que no suceda el fenómeno, o de que se elija la opción 0, se la denomina como la ODD.
- La interpretación de cada ODDS es la “ventaja” o preferencia de la opción 1 frente a la 0, es decir, el número de veces que es más probable que ocurra el fenómeno frente a que no ocurra.

La ODD, tal y como está construido (cociente entre probabilidades), siempre será mayor o igual que 0. El campo de variación de la ratio va desde 0 hasta ∞ , y su interpretación se realiza en función de que el valor sea igual, menor o superior a la unidad:

- Si **toma el valor 1** significa que la probabilidad de que ocurra la alternativa 1 es la misma que la de que no ocurra;
- Si la ratio **es menor que 1** indica que la ocurrencia de la alternativa 1 tiene menor probabilidad que la ocurrencia de la alternativa 0;
- Mientras que si **es mayor que la unidad** la opción 1 es más probable que la 0.

8. APLICACIÓN ECONOMETRICA

La metodología econométrica está formada por 4 fases: Especificación, estimación, validación y explotación del modelo. A lo largo del presente apartado vamos a ir cubriendo cada una de estas fases con el objetivo de analizar los factores que influyen sobre la probabilidad de tener una buena salud auto percibida en España.

8.1. ESPECIFICACIÓN DEL MODELO

Según la metodología econométrica, la etapa de especificación de un modelo econométrico implica la toma de varias decisiones:

- La primera decisión es la elección de la variable dependiente que presentemos explicar. En nuestro caso se trata de explicar la variable Salud. Es una variable dicotómica que toma el valor 1 si el individuo percibe una buena salud y cero en caso contrario.
- La segunda decisión importante es la elección del conjunto de las variables explicativas del modelo. En la tabla siguiente se detallan las variables explicativas elegidas. Disponemos de tres grupos de variables explicativas: Variable cualitativas de dos o más categorías; variable de escala y variables cuantitativas. En el caso de las variables cualitativas es importante especificar cuál es la categoría control sobre todo para la interpretación de los resultados de la estimación del modelo.

Tabla 9: Resumen de las variables utilizadas en el modelo

Variables explicativas	Descripción y/o Categorías	Categoría de control
Nacionalidad	Extranjero: Toma el valor 1 si es de fuera de la UE28	Español o de la UE28
Estado civil	Soltero: Toma el valor uno si es soltero	Casado
Nivel de estudio	Cualitativa de 3 categorías: E_Primeria= Estudios Hasta primarias E_Secundaria= Estudios hasta secundaria E_Superior= Tener estudios superiores	Tener estudios hasta primaria
Tipo de hogar	Cualitativa de 3 categorías: Hogar Monoparental Hogar sin hijos Hogar con hijos	Hogar con Hijos

¿Tiene alguna enfermedad o problema de salud crónicos?	E_CRONICA Toma el valor 1 si tiene enfermedad crónica	No tener enfermedad crónica
Grado de urbanización	RURAL: Toma el valor si se trata de zona poco poblada	Zona Urbana
Actividad económica	PARADO: Toma el valor 1 si está parado	Activo
Edad	Variable cuantitativa: La edad en años del individuo	
Sexo del individuo	GENERO: Toma el valor 1 si es hombre y cero en caso contrario	Mujer
FREQ_FELIZ: Frecuencia de felicidad	Variable escala de 1 a 5: 1=Siempre y 5=Nunca	
FREQ_DEP: Frecuencia con la que se sintió desanimado y deprimido	Variable escala de 1 a 5: 1=Siempre y 5=Nunca	
GS_VIDA: Grado de satisfacción global con su vida en la actualidad	Variable escala de 1 a 10: 0=Nada satisfecho y 10=Plenamente Satisfecho	

- Una vez seleccionadas las variables explicativas, el siguiente paso en la especificación del modelo econométrico es elegir la forma funcional del modelo. Teniendo en cuenta que la variable dependiente a explicar es dicotómica se puede elegir entre el modelo Logit y el Probit. Generalmente, los resultados de ambos modelos suelen ser muy parecidos. En el presente trabajo se ha optado por utilizar el modelo Logit.

8.2. ESTIMACIÓN DEL MODELO

Una vez tenemos el modelo econométrico especificación, a continuación, se estiman los diferentes modelos considerados. Para ello, utilizando el software econométrico Gretl se han estimado tres modelos Logit Binarios. El primer modelo estima la probabilidad de tener buena salud para la población en general, y a continuación se estima la misma probabilidad, pero diferenciado por sexo, hombres y mujeres.

8.3. VALIDACIÓN DE LOS MODELOS ESTIMADOS

Cada uno de los modelos estimados anteriormente ha sido sometido a una serie de contrastes de validación con el objetivo de garantizar que el modelo especificado cumple los supuestos básicos. Para ello hemos utilizado el paquete de función de Gretl

llamado BMST realizado por Artur Tarasso que sirve para los contrastes de especificación de los modelos LOGIT y PROBIT:

Ilustración 4: Función BMST de Artur Tarasso

Paquete	← Versión	Fecha	← Autor	← Resumen
a_eff	0.71	2017-11-24	Oleh Komashko	advanced treatment of marginal effects
addlist	1.2	2010-12-17	Allin Cottrell	Sequential addition of variables to a model
ADFGTest	0.055	2015-04-12	Oleh Komashko	ADF-GLS test with interpolated p-values
ADMBP	0.96	2018-05-03	Artur Tarassow	ARDL Dynamic Multiplier Bootstrap Package
almonreg	0.4	2015-02-09	Allin Cottrell	PDL (Almon lag) model
arima_sim	0.23	2018-11-12	Oleh Komashko	simulation of arima(p, d, q; P, D, Q) trajectories
armax	0.101	2017-09-22	Yi-Nung Yang	Automatically determine and show the best ARMAX model bas
BMST	1.0	2016-11-04	Artur Tarassow	Logit + Probit model specifcation tests package
BreitungCandelonTest	2.5	2018-05-29	Sven Schreiber	Breitung-Candelon test of frequency-wise Granger (non-) caus
Brown	1.5	2015-02-18	Ignacio Diaz-Emparanza	Brown linear and quadratic trend models
buys_ballot	2.0	2018-02-15	Ignacio Diaz-Emparanza,...	Plots for seasonal time series
Causa_Horizon	1.11	2018-03-31	Ignacio Diaz-Emparanza	Seasonal stability tests

Ilustración 5: Descripción de Gretl de la función BMST de Artur Tarasso

```
File: E:\Usuarios\usuario\AppData\Roaming\gretl\functions\BMST.gfn

Package: BMST 1.0 (2016-11-04)
Author: Artur Tarassow
Email: artur.tarassow@gmail.com
Required gretl version: 1.10.0
Data requirement: Algún tipo de conjunto de datos
Description: Logit + Probit model specifcation tests package

Public interfaces:

  setupBMST()
  Logit_burr()
  Logit_hom()
  Logit_reset()
  Probit_hom()
  Probit_reset()

Help text:

*** Binary Model Specification Test PACKAGE ***

This code comprises 3 specifcation tests for binary Logit models
and 2 specifcation tests for binary Probit models.
```

Ilustración 6: Detalles de la función BMST

```
set echo off
set messages off
# contrastes de especificacion del logit
include BMST.gfn
print " CONTRASTES DE ESPECIFICACIÓN DEL MODELO LOGIT"
printf ""
print "=====
printf ""
type = 1
list x = const GENERO E_CRONOCA PARADO Z_URBANA EXTRANJERO SOLTERO HOGAR_MONO HOG

logit BUENA_SALUD x
# setupBMST() must always be called 1st:
bundle b = setupBMST(BUENA_SALUD, $xlist, $yhat, $coeff, type, 1)
Logit_burr(&b) # Next, call the test of interest
print b # Print information stored in the bundle
scalar burr_pval = b.Log_burr_pval # Store p-value of Logistic Against Burr Type
print burr_pval
# Call other tests
Logit_reset(&b)
Logit_hom(&b)
print "=====
```

Los resultados de dichos contrastes realizados para los tres modelos estimados se recogen en el siguiente cuadro:

Tabla 10: Resultado de los contrastes para el modelo general

	Modelo completo	Modelo Mujeres	Modelo Hombres
Logit model LM test for correct functional form ^a :	LM-RESET = 1,185 (p-valor=0,41)	LM-RESET = 1,315 (p-valor=0,09)	LM-RESET = 1,273 (p-valor=0,10)
Logit model LM test for homoscedasticity ^b :	62,467 (p-valor=0,000)	37,926 (p-valor=0,000)	38,880 (p-valor=0,000)

a. H0: Correcta forma funcional y Ha: Incorrecta forma funcional

b. H0: Homoscedasticidad y Ha: Heteroscedasticidad

Como se puede observar, al 5% de nivel de significación los valores de contraste Reset indican que los modelos Logit están correctamente especificación, en el sentido de que presentan una correcta forma funcional. Sin embargo, los resultados indican que los modelos presentan problemas de heterocedasticidad. Para tratar de solucionar el problema y con el objetivo de poder hacer inferencias estadísticas sobre los modelos estimados, se han utilizados las desviaciones típicas robustas a la heteroscedasticidad.

Los resultados de los tres modelos finalmente estimados se recogen en las tablas 11, 12 y 13.

Tabla 11: Estimación del modelo Logit para toda la muestra (T=27714)

Variables	coeficiente	Desv. Típica	t-ratio	Pendiente (%)
Const	1,42	0,19	7,68	
GENERO	0,07	0,04	1,75	0,9%
E_CRONCA	-2,64	0,04	-66,00	-44,5%
PARADO	0,15	0,06	2,39	2,0%
Z_URBANA	0,13	0,04	3,25	1,7%
EXTRANJERO	-0,30	0,08	-3,92	-4,4%
SOLTERO	-0,10	0,06	-1,60	-1,3%
HOGAR_MONO	0,09	0,08	1,23	1,2%
HOGAR_SH	-0,04	0,05	-0,82	-0,6%
E_SECUNDARIA	0,28	0,05	5,66	3,8%
E_SUPERIOR	0,69	0,06	11,66	8,4%
EDAD	-0,04	0,00	-22,66	-0,6%
PREQ_FELIZ	-0,17	0,02	-6,96	-2,3%
GS_VIDA	0,24	0,01	18,71	3,2%
FREQ_DEPR	0,38	0,02	17,09	5,1%
<p>Media de la vble. dep. 0,718554 D.T. de la vble. dep. 0,449713</p> <p>R-cuadrado de McFadden 0,457765 R-cuadrado corregido 0,456854</p> <p>Log-verosimilitud -8931,055 Criterio de Akaike 17892,11</p> <p>Criterio de Schwarz 18015,56 Crit. de Hannan-Quinn 17931,87</p> <p>Número de casos 'correctamente predichos' = 23811 (85,9%)</p> <p>f(beta'x) en la media de las variables independientes = 0,135</p> <p>Contraste de razón de verosimilitudes: Chi-cuadrado (14) = 15079,5 [0,0000]</p>				

Tabla 12: Estimación del modelo Logit para los hombres (T=13218)

Variables	coeficiente	Desv. Típica	Z	Pendiente (%)
Const	1,37	0,27	5,05	
E_CRONCA	-2,61	0,06	-44,64	-41,0%
Z_URBANA	0,21	0,06	3,64	2,4%
EXTRANJERO	-0,31	0,12	-2,67	-4,0%
SOLTERO	-0,11	0,08	-1,26	-1,2%
HOGAR_MONO	0,06	0,11	0,57	0,7%
HOGAR_SH	0,01	0,07	0,08	0,1%
E_SECUNDARIA	0,20	0,07	2,81	2,3%
E_SUPERIOR	0,66	0,09	7,78	6,9%
PREQ_FELIZ	-0,17	0,04	-4,89	-2,0%
GS_VIDA	0,26	0,02	13,67	3,0%
FREQ_DEPR	0,39	0,03	12,15	4,6%
PARADO	0,30	0,10	3,14	3,2%
EDAD	-0,04	0,00	-16,69	-0,5%
<p>Media de la vble. dep. 0,746633 D.T. de la vble. dep. 0,434956</p> <p>R-cuadrado de McFadden 0,442619 R-cuadrado corregido 0,440748</p> <p>Log-verosimilitud -4170,007 Criterio de Akaike 8368,014</p> <p>Criterio de Schwarz 8472,865 Crit. de Hannan-Quinn 8403,019</p> <p>Número de casos 'correctamente predichos' = 11394 (86,2%)</p> <p>f(beta'x) en la media de las variables independientes = 0,116</p> <p>Contraste de razón de verosimilitudes: Chi-cuadrado (13) = 6622,86 [0,0000]</p>				

Tabla 13: Estimación del modelo Logit para las mujeres (T=14496)

Variables	coeficiente	Desv. Típica	t-ratio	Pendiente (%)
Const	1,51	0,25	6,14	
E_CRONICA	-2,67	0,05	-49,37	-47,4%
Z_URBANA	0,05	0,05	1,01	0,8%
EXTRANJERO	-0,29	0,10	-2,90	-4,8%
SOLTERO	-0,09	0,08	-1,05	-1,3%
HOGAR_MONO	0,10	0,10	1,00	1,5%
HOGAR_SH	-0,09	0,07	-1,29	-1,4%
E_SECUNDARIA	0,35	0,07	5,26	5,3%
E_SUPERIOR	0,71	0,08	8,69	9,9%
PREQ_FELIZ	-0,17	0,03	-4,99	-2,5%
GS_VIDA	0,23	0,02	13,23	3,5%
FREQ_DEPR	0,36	0,03	12,48	5,6%
PARADO	0,03	0,09	0,34	0,5%
EDAD	-0,04	0,00	-16,43	-0,6%
Media de la vble. dep. 0,692950 D.T. de la vble. dep. 0,461286 R-cuadrado de McFadden 0,468306 R-cuadrado corregido 0,466740 Log-verosimilitud -4753,331 Criterio de Akaike 9534,661 Criterio de Schwarz 9640,804 Crit. de Hannan-Quinn 9569,937 Número de casos 'correctamente predichos' = 12415 (85,6%) f(beta'x) en la media de las variables independientes = 0,153 Contraste de razón de verosimilitudes: Chi-cuadrado (13) = 8373,29 [0,0000]				

8.4. CONTRASTES DE HIPÓTESIS

En primer lugar, sobre los modelos estimados, se ha contrastado la significatividad individual de los parámetros con el objetivo de determinar cuáles son los factores estadísticamente significativos para explicar la probabilidad de tener una buena salud auto percibida:

El contraste a realizar quedaría definido como:

$H_0 : \beta = 0$ El parámetro es igual a cero

$H_1 : \beta \neq 0$ El parámetro es distinto de cero

El estadístico de contraste: $t - \text{ratio} = \frac{\hat{\beta}_i}{\sigma_{\hat{\beta}_i}} \sim N(0, 1)$

Se dice que el parámetro no es estadísticamente significativo si no se rechaza la hipótesis nula al 5% de nivel de significación (valor absoluto del estadístico t -ratio < valor crítico).

En los cuadrados las variables marcadas en color rojo son variables que han resultado estadísticamente no significativas al 5% de nivel de significación. Es decir, las que estadísticamente no contribuyen para explicar la probabilidad de tener una buena salud. Entre estos factores cabe destacar que el género, el estado civil y el tipo de hogar no influyen significativamente el estado de salud auto percibida por los individuos. En lo referente al género, hay que matizar que el hecho de que este factor no sea significativo no quiere decir que no influya en la probabilidad de tener buena salud, ya que el efecto de las variables, como se verá más adelante es diferente dependiendo del sexo. Debido a que en este tipo de modelos no se puede hacer el test Chow, se estimará un modelo para hombres y uno para mujeres.

De otro lado, las variables marcadas en azul y negro son las que han resultado ser estadísticamente significativas al 5% de significación y que, por lo tanto, sí contribuyen a explicar la probabilidad de tener buena salud. En todos los modelos aparecen como significativos los factores: si se padece o no una enfermedad crónica, la nacionalidad, el nivel educativo, la edad, el grado de satisfacción con su vida y la frecuencia con la que se tienen pensamientos tanto felices como depresivos. Además, existen dos factores que pese a contribuir a explicar la probabilidad de tener buena salud en la población general y en los hombres, no lo hacen en las mujeres. Estos factores son la situación laboral y el grado de urbanización de la zona de residencia.

8.5. BONDAD DE AJUSTE

El uso de la función de verosimilitud en la estimación hace que la bondad del ajuste en los modelos de elección discreta sea un tema controvertido, ya que en estos

modelos no existe una interpretación tan intuitiva como en el modelo de regresión clásico. A continuación, se describen los contrastes más utilizados en la literatura econométrica para medir la bondad de ajuste en un modelo Logit y que concretaremos en: índice de cociente de verosimilitudes, el estadístico chi-cuadrado de Pearson, el porcentaje de aciertos estimados en el modelo, y la prueba de Hosmer-Lemeshow.

La función de verosimilitud puede también utilizarse para obtener un estadístico, que tiene cierta semejanza con el coeficiente de determinación calculado en la estimación lineal, conocido “índice de cociente de verosimilitudes”. Este estadístico compara el valor de la función de verosimilitud de dos modelos: uno corresponde al modelo estimado que incluye todas las variables explicativas (modelo completo) y el otro sería el del modelo cuya única variable explicativa es la constante (modelo restringido). El estadístico, también conocido como R^2 de McFadden ya que fue propuesto por McFadden en 1974, se define como:

$$RV = ICV = 1 - \frac{\log L}{\log L(0)} \quad (9)$$

donde L es el valor de la función de verosimilitud del modelo completo (el estimado con todas las variables explicativas) y $L(0)$ es el valor correspondiente del modelo restringido (el que incluye únicamente en la estimación el término constante).

El ratio calculado tendrá valores comprendidos entre 0 y 1 de forma que:

- Valores próximos a 0 se obtendrán cuando $L(0)$ sea muy parecido a L , situación en la que nos encontraremos cuando las variables incluidas en el modelo sean poco significativas, es decir, la estimación de los parámetros β no mejora el error que se comete si dichos parámetros se igualaran a 0. Por lo que en este caso la capacidad explicativa del modelo será muy reducida.
- Cuanto mayor sea la capacidad explicativa del modelo, mayor será el valor de L sobre el valor de $L(0)$, y más se aproximará el ratio de verosimilitud calculado al valor 1.

Otra de las vías utilizadas para determinar la bondad de un modelo Logit es predecir con el modelo los valores de la variable endógena Y_i de tal manera que $Y_i = 1$

si $\hat{M}_i > c$ ó $Y_i = 0$ si $\hat{M}_i < c$. Generalmente, el valor que se asigna a c para determinar si el valor de la predicción es igual a 1 o a 0 es de 0,5, puesto que parece lógico que la predicción sea 1 cuando el modelo dice que es más probable obtener un 1 que un 0.

Una vez seleccionado el nivel del umbral, y dado que los valores reales de Y_i son conocidos, basta con contabilizar el porcentaje de aciertos para decir si la bondad del ajuste es elevada o no.

Una vez realizada la estimación, se puede observar el criterio de aciertos y errores, que analiza la bondad del ajuste, es decir, determina si el modelo es adecuado para explicar el comportamiento de la salud auto percibida a través de las variables exógenas introducidas. De manera general, aquel individuo para el cual se estime una probabilidad \hat{p}_i mayor o igual a 0,5 se clasificará en el grupo de aciertos ($Y_i=1$).

En este caso, el modelo general predice de manera correcta un 85,9% de los casos, lo que implica una muy buena bondad de ajuste, esto es lo mismo que indica el R^2 de McFadden, que con un valor mayor de 0,4 (0,457765) nos indica un buen ajuste tratándose de una regresión logística.

En el modelo de hombres el ajuste es ligeramente mejor que en el de mujeres respecto a los datos predichos: 86,2% en el caso de los varones y 85,6% en el caso de las mujeres y ambos coeficientes R^2 de McFadden superan el valor 0,4, lo que también indica un muy buen ajuste en este tipo de regresión.

Es por esto por lo que se puede afirmar que los tres modelos tienen una buena capacidad predictiva y que las conclusiones resultantes de los mismos se asemejan a la realidad.

8.6. INTERPRETACIÓN DE LOS RESULTADOS

En las tablas 11-13 se muestran tanto los parámetros estimados como las pendientes correspondientes que, en realidad miden el efecto marginal de cada variable explicativa sobre la probabilidad de tener buena salud.

En este caso, de entre las variables que sí son significativas para el modelo, se pueden encontrar dos tipos: aquellas que afectan positivamente a la variable endógena y las que lo hacen de manera negativa.

Las variables cuyo efecto sobre la variable dependiente es positivo, aparecen en la tabla en negro y las que influyen de manera negativa (es decir, reducen las probabilidades de tener buena salud) aparecen en color azul.

Resulta curioso ver que el hecho de ser parado aumenta la probabilidad de tener buena salud. En concreto, un parado tendría un 2% más de probabilidades que los activos de reflejar un buen nivel de salud. Aunque cabría esperar lo contrario, la precariedad laboral, por ejemplo, podría ser uno de los factores que resulten determinantes en este efecto, aunque sería interesante ahondar en el estudio de la relación entre estas dos variables.

Respecto al lugar de residencia, según los resultados que se pueden observar en el cuadro, las personas que vivan en entornos urbanos tienen un 1,7% más de probabilidades de tener buena salud que quienes vivan en entornos rurales, este efecto resulta influido por el papel que juega la edad en el nivel de salud, ya que quienes viven en entornos menos poblados suelen ser personas mayores. Como se puede ver en el cuadro, la edad supone una probabilidad de detrimento de la salud de un 0,6% por cada año que una persona envejece, por lo que, en el caso de las personas mayores (mayores de 65 años), su salud se ve enormemente influida de manera negativa por su edad.

Además, el factor estudios influye de una manera importante, de tal manera que aquellas personas cuyo máximo grado de educación cursada sea secundaria tienen un 3,8% más de probabilidades de tener buena salud, porcentaje que aumenta al 8,4% en el caso de los estudios superiores.

Esta evidencia muestra que la educación es un factor que influye de manera determinante en la salud y que, un mayor nivel de estudios cursados significa normalmente un mejor estado de salud.

Al aumentar el grado de satisfacción con la vida, aumenta la probabilidad de tener buena salud en un 3,2% por cada punto que incrementa su satisfacción. En la misma línea, pero con un efecto aún más importante encontramos la frecuencia con la que se

tienen pensamientos depresivos. A nivel que los pensamientos tristes o deprimentes se van haciendo menos recurrentes o su fuerza disminuye, la probabilidad de que el individuo tenga buena salud aumenta en un 5,1%. Estos dos efectos, junto con el efecto negativo que se observa en la variable que indica la frecuencia con la que se tienen pensamientos felices (la disminución de un nivel en la frecuencia con la que aparecen estos pensamientos en la mente de una persona, reduce su probabilidad de tener buena salud en un 2,3%) indican, que el optimismo (que finalmente es a lo que se refieren estas tres variables) es un factor determinante para el estado de la salud auto percibida. No resulta descabellado pensar que aquellas personas acostumbradas a ver la vida de una manera más optimista tendrán el mismo punto de vista en lo que respecta a su salud.

La nacionalidad también influye de manera significativa y negativa en la salud, de manera que las personas con una nacionalidad de un país que no sea europeo tienen un 4,4% de probabilidades de tener peor salud que alguien con nacionalidad europea.

Por último, el efecto más importante lo encontramos en lo que se refiere a padecimiento de enfermedades crónicas. Sufrir este tipo de problemas incrementa la probabilidad de tener mala salud en un 44,5% frente a alguien libre de este tipo de patologías.

Tabla 14: Comparación de los resultados entre hombres y mujeres

Variables	Hombres	Mujeres	Estadístico t	¿Dif sig al 5% ?
Const				
E_CRONICA	-41,0%	-47,4%	-6,903	SI
Z_URBANA	2,4%	0,8%	-1,931	SI
EXTRANJERO	-4,0%	-4,8%	-0,485	NO
SOLTERO	-1,2%	-1,3%	-0,059	NO
HOGAR_MONO	0,7%	1,5%	0,508	NO
HOGAR_SH	0,1%	-1,4%	-1,350	NO
E_SECUNDARIA	2,3%	5,3%	2,945	SI
E_SUPERIOR	6,9%	9,9%	2,815	SI
PREQ_FELIZ	-2,0%	-2,5%	-1,057	NO
GS_VIDA	3,0%	3,5%	1,854	NO
FREQ_DEPR	4,6%	5,6%	2,248	SI
PARADO	3,2%	0,5%	-2,047	SI
EDAD	-0,5%	-0,6%	-2,888	SI

Tabla 15: Comparación de los ODD-Ratio y sus IC entre el modelo de hombres y el de mujeres

Variable	Hombres		Mujeres	
	Odds-ratio	95,0% conf. Interval	Odds-ratio	95,0% conf. Interval
E_CRONOCA	0,0735	[0,066, 0,082]	0,0692	[0,062, 0,077]
Z_URBANA	1,2313	[1,101, 1,378]	1,0556	[0,950, 1,172]
EXTRANJERO	0,7332	[0,584, 0,921]	0,746	[0,612, 0,910]
SOLTERO	0,8997	[0,763, 1,060]	0,9185	[0,784, 1,076]
HOGAR_MONO	1,0664	[0,854, 1,331]	1,1024	[0,910, 1,335]
HOGAR_SH	1,0056	[0,871, 1,161]	0,9146	[0,799, 1,047]
E_SECUNDARIA	1,2229	[1,063, 1,407]	1,4202	[1,246, 1,619]
E_SUPERIOR	1,9401	[1,642, 2,293]	2,0328	[1,732, 2,386]
PREQ_FELIZ	0,8414	[0,785, 0,902]	0,8468	[0,793, 0,904]
GS_VIDA	1,293	[1,246, 1,341]	1,2552	[1,214, 1,298]
FREQ_DEPR	1,4822	[1,391, 1,579]	1,4399	[1,360, 1,525]
PARADO	1,3544	[1,121, 1,637]	1,0308	[0,865, 1,229]
EDAD	0,9569	[0,952, 0,962]	0,9603	[0,956, 0,965]

Entre los modelos de hombres y mujeres se pueden observar claras diferencias que se manifiestan de dos formas. Por un lado, las variables que no son estadísticamente significativas no son las mismas para cada grupo. A las variables no relevantes comentadas anteriormente, en el caso de las mujeres habría que sumar el nivel de población de la zona de residencia y la actividad económica que desempeñen (si son trabajadores parados o activos). Por otro lado, las diferencias entre hombres y mujeres se manifiesta en el efecto marginal de cada variables sobre la probabilidad de tener una buena salud autopercebida. Para poner de manifiesto este hecho se ha realizado el siguiente contraste:

$$H_0 : \beta_i^{\text{Mujer}} = \hat{\beta}_i^{\text{Hombre}}$$

$$H_a : \beta_i^{\text{Mujer}} \neq \hat{\beta}_i^{\text{Hombre}}$$

El estadístico de contraste es:

$$t = \frac{\hat{\beta}_i^{\text{Mujer}} - \hat{\beta}_i^{\text{Hombre}}}{\hat{\sigma}_{\hat{\beta}_i^{\text{Mujer}}}} \sim N(0,1)$$

Rechazar la hipótesis nula implica que el efecto de la variable X_i es estadísticamente diferente según si trata de hombre o una mujer. Los resultados de dicho contraste aparecen recogidos en la Tabla 14. Como se puede apreciar hay diferentes variables que tienen diferentes efectos sobre la probabilidad de tener buena salud, entre las cuales cabe destacar el nivel de educación. El hecho de tener nivel de educación superior aumenta la probabilidad de tener una buena salud en un 10% en el caso de las mujeres y solo el 7% en el caso de los hombres.

En el caso de los hombres, vivir en una zona urbana, hace que tengas 1,23 veces más probabilidades de tener buena salud que si habitase una zona rural. Aún más fuerte es la relación que se establece entre estar parado y tener buena salud. Esta variable, que para las mujeres no es significativa, supone que los hombres parados tengan 1,4 veces más probabilidades de tener buena salud que quienes tienen una situación laboral activa.

Aparte de esto, como vemos en el cuadro: de comparación de las pendientes entre hombres y mujeres, Las diferencias entre los distintos efectos que causan las variables en la endógena dependiendo del sexo pueden no ser significativas, es decir, que estas diferencias entre hombres y mujeres no son significativas respecto de la manera en que afectan a la buena salud.

Como podemos ver, Existen varios factores que afectan de una manera más pronunciada a mujeres que a hombres. El hecho de padecer una patología crónica afecta más negativamente a las ellas (su probabilidad de tener mala salud aumenta un 47,4%) que a ellos (cuya probabilidad de tener mala salud aumenta un 41%).

Aunque entre las variables que hacen referencia a la frecuencia con la que se tienen pensamientos felices y al grado de satisfacción con su vida no ofrecen diferencias significativas entre sexos, sí lo hace la variable que tiene que ver con pensamientos depresivos, que afectan más negativamente a mujeres (que disminuyen la probabilidad de tener buena salud en un punto porcentual más de lo que lo hacen los hombres por cada grado de frecuencia que aumenten estos pensamientos).

No sólo afectan más a mujeres que a hombres factores que influyen de manera negativa, sino que en el caso de que una mujer tenga estudios secundarios o superiores, su probabilidad de tener buena salud se incrementa un 3% más de lo que lo haría en el caso de un hombre. De hecho, según se puede apreciar en la tabla que hace referencia a los ODD-Ratio, Podemos ver cómo una mujer que tenga estudios superiores tendrá 2,03 veces más probabilidades de tener buena salud que una mujer con estudios primarios. En el caso de los hombres, quienes tengan estudios de más alto grado, tendrán 1,94 veces más probabilidades de tener buena salud. Es decir, la paridad en ambos casos rondará el 2:1, lo que significa que tener estudios superiores prácticamente dobla las posibilidades de que una persona refiera buena salud.

9. CONCLUSIONES

La salud auto percibida es un importante indicador de cara a la evaluación y diseño de políticas sociales, económicas y de salud, ya que son múltiples los factores que influyen en ella.

En 2018, la población española de acuerdo con los resultados de la Encuesta de Condiciones de Vida del INE reflejó cómo ciertos determinantes sociales de la salud como el tipo de hogar en el que se vive, la situación respecto a la actividad productiva de las personas, el sexo o la edad influyen de manera importante en este indicador de salud general.

Uno de los resultados llamativos obtenido en este trabajo es que los parados tienen un mejor estado de salud auto percibido que las personas laboralmente activas.

Los modelos que se han estimado corroboran que la salud auto percibida empeora a medida que aumenta la edad y que, además, esta relación entre la mayor edad y una peor salud es más fuerte en el caso de las mujeres que en el de los hombres.

También se ha observado que el nivel máximo de estudios cursados influye de una manera más fuerte en mujeres que en hombres, a mayor nivel de estudios, mejor salud, incrementando la probabilidad de disfrutar de una buena salud de manera más notable

aún si la persona tiene o cursa estudios superiores frente a quienes tienen estudios primarios.

También afecta en menor medida a los varones el hecho de padecer enfermedades crónicas, aunque este es el factor que sin duda más influye en la salud de los que se tratan en el estudio. Este tipo de enfermedades suelen influir de manera importante en el día a día de quienes las padecen, por lo que es difícil que quienes las sufran perciban un buen estado de salud.

10. BIBLIOGRAFÍA

AGUILAR-PALACIO, I. P. CARRERA-LADFUENTES & M.J. RABANAQUE. 2014. "Salud percibida y nivel educativo en España: tendencias por comunidades autónomas y sexo (2001-2012)" *Gaceta Sanitaria* 2015. (núm. 29 (1), pág 37-43)

ÁLVAREZ B. 2007. "Modelos de elección binaria" Publicación de la asignatura Econometría II de la Universidad de Vigo.

ARTAZCOZ L, ARTIEDA L, BORRELL C, ET AL. 2004. "Combining job and family demands and being healthy: What are the differences between men and women?". *Eur J Public Health*, 2004 (núm14, pág 43-48)

DAHLGREN, G. & WHITEHEAD, M. 2006. "Concepts and principles for tackling social inequities in health: Levelling up Part 2." Publicación de la Oficina Regional de la OMS para Europa. 2006.

FRENZ, P. 2005. "Desafíos en salud pública de la Reforma, equidad y determinantes sociales de la salud", *Revista de Salud Pública de Chile* (Vol 9 (2), pág. 103-110). Chile.

GIRÓN P. 2010. "Los determinantes de la salud percibida en España". Tesis Doctoral, Universidad Complutense de Madrid; 2010.

GUMÀ, J. B. ARPINO & A.SOLÉ-AURÓ. 2017. "Determinantes sociales de la salud de distintos niveles por género: educación y hogar en España" *Gaceta Sanitaria* 2019. (núm. 33 (2), pág 127-133)

GUMÀ, J. & R. TREVIÑO. 2015. "Posición en el hogar y género. Desigualdades en la calidad de vida relacionada con la salud entre la población adulta en España", *Revista Internacional de Sociología*, 73, <http://dx.doi.org/10.3989/2013.03.04>

MORCILLO, V. 2017. "Desigualdades en la salud autopercebida de la población española. Revisión sistemática de literatura" Tesis doctoral. Facultad de Medicina, Universidad Autónoma de Madrid.

OMS. 1948. "Official Records of the World Health Organization", N° 2, p. 100. Preámbulo de la Constitución de la Organización Mundial de la Salud, adoptada por la Conferencia Sanitaria Internacional en 1946.

ROHLFS, I. C.BORRELL, C. ANITUA, L. ET AL. 2000. "La importancia de la perspectiva de género en las encuesta de salud" Gaceta Sanitaria (núm. 14 (2), pág 146-155)

SÁNCHEZ, R. 2017. "La salud autopercebida, fenómeno multidimensional de gran interés en el campo de la investigación en salud" Blog Nuestros expertos, Universidad Internacional de Valencia. Ciencias de la salud.

VERBRUGGE LM. 1989. "The twain meet: Empirical explanations of sex differences in health and mortality". J Health Soc Behav. 1989, núm30, págs282-304.

